Ai Waifu Rat explora os usuários com táticas avançadas de engenharia so

Data: 2025-08-31 15:01:09

Autor: Inteligência Against Invaders

Uma nova campanha sofisticada de malware emergiu que armazia a inteligência artificial e a

engenharia social para direcionar comunidades on -line de nicho.

Os pesquisadores de segurança identificaram o “Ai Waifu Rat”, um Trojan de acesso remoto que se disfarça como uma ferramenta de interação inovadora de IA, fornecendo aos atacantes acesso completo ao sistema aos computadores das vítimas.

O malware tem como alvo especificamente Modelo de linguagem grande (LLM) Comunidades de interpretação de papéis, explorando o entusiasmo dos usuários pela tecnologia de IA de ponta e sua confiança em colegas de comunidade.

Em vez de depender puramente da sofisticação técnica, essa ameaça demonstra como os cibercriminosos modernos estão cada vez mais alavancando a manipulação psicológica para ignorar as defesas de segurança.

A campanha de ratos AI Waifu representa uma masterclass em marketing enganoso e manipulação social. O ator de ameaças, operando sob pseudônimos, incluindo Kazepsi e Psioniczephyr, se apresentou como um legítimo “jogador de criptografia da CTF” e pesquisador explorando os limites da IA.

Eles comercializaram seu software malicioso como uma emocionante “meta experiência” que permitiria aos personagens da IA ??”quebrar a quarta parede” e interagir diretamente com os computadores do mundo real dos usuários.

Táticas enganosas principais empregadas pelo ator de ameaças:

- **Credenciais falsas** – alegou ser um jogador experiente da CTF, apesar de não ter um histórico verificável da competição.
- **Apresenta reflexão** – Apresentou a execução do código arbitrário perigoso como um emocionante “recurso avançado”.
- **Infiltração da comunidade** -Construiu confiança participando das comunidades de interpretação de papéis da Niche LLM ao longo do tempo.
- **Legitimidade técnica** – usou jargão de programação e referências para criar uma aparência de especialização.

Os recursos prometidos incluíam permitir que os caracteres da IA ??leiam os arquivos locais para recursos de “interpretação de papéis personalizados” e diretos de “execução de código arbitrário”, apresentados como recursos avançados, em vez de vulnerabilidades de segurança.

Esse enquadramento mostrou -se devastadoramente eficaz na comunidade -alvo, onde os membros já estavam interessados ??em novas interações de IA e dispostas a experimentar novas tecnologias.

O invasor instruiu explicitamente os usuários a desativar software antivírus Ou adicione o binário malicioso às listas de exclusão, alegando que elas eram “falsos positivos” devido às “operações de baixo nível” do programa.

Essa tática clássica de engenharia social explorou a curiosidade técnica do público -alvo enquanto desmontava sua principal linha de defesa contra a detecção de malware.

A arquitetura técnica revela verdadeira intenção

Sob a atraente fachada de marketing, há um acesso remoto direto, mas perigoso, Trojan. O malware opera executando um agente local nas máquinas das vítimas que escuta os comandos na porta 9999.

Esses comandos, supostamente originários das interações IA, são transmitidos como solicitações HTTP de texto simples e executados diretamente no sistema de destino.

O rato expõe três pontos de extremidade críticos que fornecem acesso abrangente ao sistema. Os terminais “/execute_trusted” geram processos de PowerShell para executar comandos arbitrários, enquanto o terminal “/readfile” permite que os invasores acessem e exfilem qualquer arquivo no sistema local.

Um terceiro endpoint, “/Execute”, inclui o que parece ser um mecanismo de consentimento do usuário, mas isso prova ser um mero teatro de segurança, pois os invasores podem simplesmente ignorá -lo usando o terminal irrestrito “/execute_trusted”.

Essa arquitetura cria vários vetores de ataque além do controle do ator de ameaças originais. A comunicação HTTP de texto simples torna o sistema vulnerável a ataques de homem no meio de outros softwares maliciosos, enquanto a porta local fixa permite que sites maliciosos sequestrem potencialmente a conexão através de ataques baseados em navegador.

Padrão de comportamento malicioso e táticas de evasão

A investigação sobre a história do ator de ameaças revela um padrão consistente de práticas perigosas de programação e intenção maliciosa.

Os lançamentos anteriores incluíram cartões de caráter de AI baseados na Web que usaram funções JavaScript Eval () para executar o código gerado por LLM diretamente nos navegadores-um anti-padronização fundamental de segurança que demonstra intenção maliciosa ou profunda negligência de segurança.

Um suposto “desafio da CTF” lançado Pelo mesmo ator continha lógica explicitamente maliciosa, incluindo código que fecharia à força os computadores dos usuários se eles inserissem respostas incorretas.

O programa também implementou mecanismos de persistência e técnicas de anti-análise típicas do malware, apesar de serem comercializadas como um quebra-cabeça legítimo.

Quando os pesquisadores de segurança relataram o malware para os provedores de hospedagem, o ator de ameaças imediatamente iniciou as manobras de evasão.

Eles migraram o malware em várias plataformas, incluindo Github, Gitgud, OneDrive e Mega.NZ, frequentemente usando arquivos protegidos por senha para evitar a detecção.

O ator também criou vários aliases e contas para contornar os esforços de remoção, demonstrando uma consciência clara de suas atividades maliciosas.

A investigação revelou que, apesar das reivindicações de ser um experiente “jogador de criptografia da CTF”, não existem registros do ator de ameaças que participam da captura legítima das competições de bandeira ou das comunidades de pesquisa de segurança.

Essa falsa credencial parece fazer parte da campanha de engenharia social mais ampla, projetada para estabelecer credibilidade nas comunidades técnicas.

O incidente de rato da AI Waifu destaca um cenário emergente de ameaças, onde os cibercriminosos exploram o entusiasmo pela tecnologia de IA e confiança da comunidade para distribuir malware.

À medida que as ferramentas de IA se tornam mais integradas à computação diária, a conscientização da segurança deve evoluir para reconhecer quando “recursos inovadores” cruzam a linha em vulnerabilidades perigosas.

Indicadores de compromisso (IOCs)

Tipo de indicador	Detalhes
Hashes de arquivo (SHA256)	F64DBD93CB5032A2C89CFAF324340349BA4B D4B0EB0325D4786874667100260 7C3088F536484EAA91141FF0C10DA788240F88 73AE53AB51E1C770CF66C04B45 CDA5ECF4DB9104B5AC92B998FF60128EDA69 C2ACAB3860A045D8E747B6B5A577 6E0EA9D2FC8040CE22265A594D7DA03149875 83C0F892C67E731947B97D3C673 11B07EF15945D2F1E7CF192E49CBF670824135 562C9B87C20EBD630246AD1731 FDF461A6BD7E806B45303E3D7A76B5916A452 9DF2F4DFF830238473C616AC6F9
Nomes de arquivos	js_windows_executor.exe nulla_re.exe android_server.py
Indicadores de rede	Tráfego HTTP para 127.0.0.1:9999 do processo do agente
Persistência	Chave do Registro: HKCU Software Microsoft Windows CurrentVersion Run Nome do valor: FakeUpDater
URLs do provedor de hospedagem	https://gitgud.io/kazepsi/file-storage/-/raw/master/nulla/ctf/nulla_re.exe (já queda)

Tipo de indicador	Detalhes
	https://gitgud.io/kazepsi/file-storage/-/raw/master/backends/js_windows_executor.exe (já queda) https://gitgud.io/kazepsi/file-storage/-/raw/master/backends/android_server.py (já queda) https://github.com/pioniczephyr/files/blob/main/ctf-puzzles.json (já queda) https://github.com/pioniczephyr/files/blob/main/code/js_windows_executor.exe (já queda) https://github.com/pioniczephyr/files/blob/main/code/android_server.py (já queda) https://github.com/kazepsi/file-storage/blob/main/code/code.rar (já queda) https://1drv.ms/u/c/6b4c603601e43e48/exwj4vbq2mhiqczx6weka-abfuwr_8setpkh5k_83czhqg?e=blzzl6 (já retire) https://mega.nz/file/gfkrsaba#dmedscmvpvgf7ypum0h96ay4nbq7oe6sgzj9hq4rpk0 (já queda) https://mega.nz/file/wz9xcrbc#0mxn1gwijb41bxbvqc-bf_avpomjdbo9jk04572oih8 (queda pendente)

Encontre esta notícia interessante! Siga -nos Google News**Assim,**LinkedIn**e**X**Para obter atualizações instantâneas!**